

# 3-Keto-5-aminohexanoate Cleavage Enzyme

## A COMMON FOLD FOR AN UNCOMMON CLAISEN-TYPE CONDENSATION<sup>\*†</sup>

Received for publication, April 20, 2011, and in revised form, May 17, 2011. Published, JBC Papers in Press, June 1, 2011, DOI 10.1074/jbc.M111.253260

Marco Bellinzoni<sup>‡1</sup>, Karine Bastard<sup>§||1</sup>, Alain Perret<sup>§||</sup>, Anne Zapparucha<sup>§||2</sup>, Nadia Perchat<sup>§||</sup>, Carine Vergne<sup>§||</sup>, Tristan Wagner<sup>‡</sup>, Raquel C. de Melo-Minardi<sup>§||3</sup>, François Artiguenave<sup>§||4</sup>, Georges N. Cohen<sup>\*\*</sup>, Jean Weissenbach<sup>§||</sup>, Marcel Salanoubat<sup>§||</sup>, and Pedro M. Alzari<sup>‡5</sup>

From the <sup>‡</sup>Unité de Microbiologie Structurale, Institut Pasteur, and CNRS-URA2185, 25 rue du Dr. Roux, 75724 Paris Cedex 15, the <sup>§</sup>Direction des Sciences du Vivant, Commissariat à l'Energie Atomique (CEA), Institut de Génomique, Genoscope, 2 rue Gaston Crémieux, 91057 Evry, <sup>||</sup>CNRS-UMR8030, 2 rue Gaston Crémieux, 91057 Evry, the <sup>||</sup>Université d'Evry Val d'Essonne, boulevard François Mitterrand, 91057 Evry, and the <sup>\*\*</sup>Institut Pasteur, 28 rue du Dr. Roux, 75724 Paris Cedex 15, France

The exponential increase in genome sequencing output has led to the accumulation of thousands of predicted genes lacking a proper functional annotation. Among this mass of hypothetical proteins, enzymes catalyzing new reactions or using novel ways to catalyze already known reactions might still wait to be identified. Here, we provide a structural and biochemical characterization of the 3-keto-5-aminohexanoate cleavage enzyme (Kce), an enzymatic activity long known as being involved in the anaerobic fermentation of lysine but whose catalytic mechanism has remained elusive so far. Although the enzyme shows the ubiquitous triose phosphate isomerase (TIM) barrel fold and a Zn<sup>2+</sup> cation reminiscent of metal-dependent class II aldolases, our results based on a combination of x-ray snapshots and molecular modeling point to an unprecedented mechanism that proceeds through deprotonation of the 3-keto-5-aminohexanoate substrate, nucleophilic addition onto an incoming acetyl-CoA, intramolecular transfer of the CoA moiety, and final retro-Claisen reaction leading to acetoacetate and 3-aminobutyl-CoA. This model also accounts for earlier observations showing the origin of carbon atoms in the products, as well as the absence of detection of any covalent acyl-enzyme intermediate. Kce is the first representative of a large family of prokaryotic hypothetical proteins, currently annotated as the “domain of unknown function” DUF849.

One of the most striking outputs of genome sequencing is the high proportion of proteins without pertinent functional annotation. Predicted genes have been mainly annotated on the basis of sequence homology to already characterized proteins from other genomes, resulting, on average, in a fraction as high as 30–50% of predicted gene products that, in a given genome, have no accurate functional assignments. These hundreds of proteins without known function within a single organism prevent a deeper understanding of the cell as a biological system because it might be expected that some of these proteins of yet undiscovered functionalities may end up revealing significant novelty in chemical reactions.

The increasing numbers of proteins of unknown function from newly sequenced genomes have been grouped into families and classified within the Pfam database as “domain of unknown function” (DUF)<sup>6</sup> (1). Besides grouping protein families by sequence similarities (2, 3) and gene context analyses (4), structural genomics has emerged as a promising tool for functional annotation, given the assumption that proteins with a similar fold can be related in function (5). To narrow down the possible functions of each family and to achieve a better perception of the protein universe, the NIH Protein Structure Initiative (PSI; [www.nigms.nih.gov/Research/FeaturedPrograms/PSI](http://www.nigms.nih.gov/Research/FeaturedPrograms/PSI)) has made a systematic effort to produce representative structures of most DUF families (6). Our attention was drawn by DUF849, a family restricted to prokaryotes and found in ~850 bacterial proteins belonging to several different phyla where, according to gene predictions, it accounts in most (albeit not all) cases for the whole protein chain. Despite the fact that a few crystallographic structures of proteins associated with DUF849 have been recently deposited in the Protein Data Bank (PDB), no mechanistic or biochemical investigation is at present available. In a previous work, we identified a gene associated with DUF849 as coding for the 3-keto-5-aminohexanoate cleavage enzyme (Kce) involved in the lysine fermentation pathway (7), an intriguing enzyme whose catalytic mechanism has remained so far enigmatic. The reaction catalyzed by Kce involves the reversible condensation of a six-carbon molecule, 3-keto-5-aminohexanoate (KAH), derived from L-lysine,

<sup>\*</sup> This work was supported by grants from the Institut Pasteur, the CEA, the CNRS, and the University of Evry.

<sup>†</sup> This article was selected as a Paper of the Week.

The atomic coordinates and structure factors (codes 2Y7D, 2Y7E, 2Y7F, and 2Y7G) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

<sup>‡</sup> The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Experimental Procedures, Tables S1–S3, and Figs. S1–S3.

<sup>1</sup> Both authors contributed equally to this work.

<sup>2</sup> To whom correspondence may be addressed: Laboratoire de Chimie Organique et Biocatalyse, IG Genoscope and CEA-UMR8030, 2 rue Gaston Crémieux, 91057 Evry cedex, France. Fax: 33-1-60872514; E-mail: [azaparuc@genoscope.cns.fr](mailto:azaparuc@genoscope.cns.fr).

<sup>3</sup> Present address: Dept. of Computer Science, Federal University of Minas Gerais, 6627 Pampulha, Belo Horizonte, Brazil.

<sup>4</sup> Present address: Medical Genetics Dept., School of Medical Sciences, University of Campinas, UNICAMP, Campinas, São Paulo 13083-000, Brazil.

<sup>5</sup> To whom correspondence may be addressed. Fax: 33-1-45688604; E-mail: [pedro.alzari@pasteur.fr](mailto:pedro.alzari@pasteur.fr).

<sup>6</sup> The abbreviations used are: DUF, domain of unknown function; KAH, 3-keto-5-aminohexanoate; Kce, 3-keto-5-aminohexanoate cleavage enzyme; TIM, triose phosphate isomerase; r.m.s., root mean square; MAD, multiwavelength anomalous diffraction.

through three different reactions with acetyl-CoA to produce two four-carbon species, of which one still esterified to CoA, *i.e.* acetoacetate and a 3-aminobutyryl-CoA (8–10) (see Fig. 1A). Such  $\beta$ -keto acid-degrading activity proceeding through the cleavage on the  $\beta$ -carbon still has no analogues in literature and therefore makes the characterization of this enzyme of particular interest.

Here, we investigated the catalytic mechanism of Kce from *Candidatus* Cloacamonas acidaminovorans by structural and biochemical means. We determined the crystal structures of the enzyme in its native state, as well as complexed with either one of its natural substrates ((S)-KAH) or one of the products (acetoacetate). Additionally, we modeled the assumed intermediates of the reaction into the binding pocket, resulting in poses that agree with the proposed catalytic mechanism. The ensemble of our results suggests that the enzyme, although structurally similar to class II aldolases, proceeds through a Claisen-like condensation with the peculiarity that CoA moiety is transferred in an intramolecular fashion to the keto group of the KAH moiety.

### EXPERIMENTAL PROCEDURES

Experimental procedures for DNA cloning, protein production, purification and crystallization, biochemical assays, and site-directed mutagenesis are described in the [supplemental material](#). Oligonucleotide sequences are provided in [supplemental Table S1](#).

**Data Collection, Structure Solution, and Refinement**—X-ray diffraction data were collected either at Synchrotron Soleil (Saint-Aubin, France) or at the European Synchrotron Radiation Facility (ESRF) (Grenoble, France) from single crystals at 100 K. Diffraction frames were processed with XDS (11) followed by data reduction with either XSCALE (11) or SCALA within the CCP4 suite of programs (12) (see Table 1 and [supplemental Table S1](#)). The structure was solved by multiwavelength anomalous diffraction (MAD) phasing from Se-Met-labeled protein crystals. A total of 27 selenium sites in the asymmetric unit were located with the program SHELXD (13) on data from a redundant single anomalous dispersion dataset from a Se-Met-labeled crystal ([supplemental Table S2](#)); three-wavelength MAD phasing and density modification were then carried out on data collected from an isomorphous crystal with SOLVE/RESOLVE (14). The resulting electron density map showed clear secondary structure allowing the manual tracing of the polypeptide chain with Coot (15). This preliminary model was partially refined with Refmac5 (16) and then used to perform molecular replacement on higher resolution data from a different crystal form (space group  $P2_12_12_1$ ). All subsequent complexes with ligands were solved by molecular replacement with the program Molrep (17) using the structure of the Kce monomer, without ligands, as the search model. Each model was inspected and rebuilt with Coot and refined with autoBUSTER (18). All models were validated through the Molprobity server (19). Final refinement parameters are shown in Table 1. Figures were generated and rendered with the PyMOL Molecular Graphics System, Version 1.3, (Schrödinger, LLC); electrostatic surfaces were rendered with the APBS plug-in within PyMOL, supplying atomic charges and radii calculated

through the Poisson-Boltzmann pdb2pqr server (20) applying the CHARMM force field.

**Substrate Synthesis**—(S)-KAH and (R)-KAH were synthesized from (*S,S*)-(+)-methyl *N*-(*p*-toluenesulfinyl)-3-amino-3-methylpropanoate (21) and (*R,R*)-(–)-methyl *N*-(*p*-toluenesulfinyl)-3-amino-3-methylpropanoate, respectively. Detailed experimental procedures and compound characterization data, including  $^1\text{H}$  and  $^{13}\text{C}$  NMR data, are available in the [supplemental material](#).

**Molecular Modeling**—Acetyl-CoA and 3-aminobutyryl-CoA chemical depictions were retrieved from the PubChem Compound database ([pubchem.ncbi.nlm.nih.gov](#)). Intermediates of reaction were built with ChemDraw, whereas docking simulations were run with Autodock Vina (22). Energy minimization of intermediate I1 (the deprotonated form of KAH) and acetyl-CoA in the binding pocket proceeded using MMTK (23) within the Chimera package (24). More details of the computational methodology are provided in the [supplemental material](#).

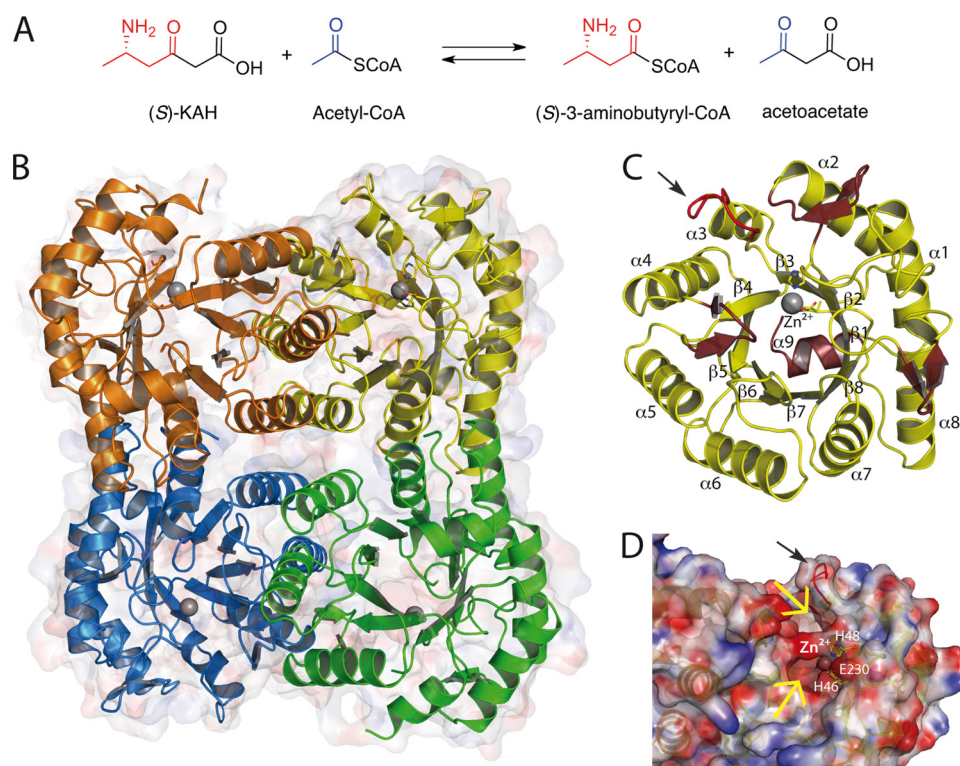
### RESULTS AND DISCUSSION

**Overall Structure**—The structure of Kce was solved by multiwavelength anomalous diffraction on selenomethionyl-labeled protein crystals diffracting at 3.2 Å resolution ([supplemental Table S2](#)). The non-isomorphous structures of the native enzyme at higher resolution (1.6 Å), as well as the complexes with substrates and products, were solved by molecular replacement; a total of four separate models were refined at resolutions ranging from 1.75 to 1.28 Å (Table 1). In all the crystal structures determined, the enzyme is assembled as a homotetramer with point symmetry 222 (Fig. 1B), consistent with the observed tetrameric state of the recombinant enzyme in solution (7) and with earlier observations on purified bacterial homologs from other species (8–10). Each subunit, made of 276 residues, shows the canonical ( $\beta/\alpha$ )<sub>8</sub> TIM barrel fold, with a short additional C-terminal  $\alpha$ -helix ( $\alpha_9$ ) protruding at the N-terminal face of the barrel and three  $\beta$ -turn extensions coming out from the barrel core, inserted in between  $\beta_2$  and  $\alpha_2$  (residues 49–58),  $\beta_4$  and  $\alpha_4$  (residues 112–119), and  $\beta_8$  and  $\alpha_8$  (residues 234–240) (Fig. 1C). The only missing part in the model is the loop connecting  $\beta_3$  to  $\alpha_3$  (residues 85–92), due to the lack of supporting electron density in two out of the four monomers, as a consequence of its high conformational flexibility (Fig. 1B). Each protomer buries a total of  $\sim 2100 \text{ Å}^2$  (18%) of the monomer surface in intersubunit contacts, which mainly involve helices  $\alpha_5$  and  $\alpha_6$ , as well as the  $\beta$ -turn  $\beta_{4a}$ – $\beta_{4b}$ , for one dyad, and  $\alpha_7$  and  $\alpha_8$  for the other dyad. As observed in the vast majority of TIM barrel enzymes, the active site is situated in each monomer at the C-terminal face of the central  $\beta$ -barrel and could be identified by the presence of a metal ion coordinated by His-46, His-48, and Glu-230 (Fig. 1D). This ion was modeled as  $\text{Zn}^{2+}$  according to its coordination geometry, and its nature was then confirmed by the analysis of double difference anomalous maps calculated from data collected around the zinc K-edge ([supplemental Fig. S1A](#) and [supplemental Table S3](#)). The metal ion is situated at the bottom of a crevice accessible from two opposite openings, one of which is delimited by the mobile  $\beta_3$ – $\alpha_3$  loop (Fig. 1D). A structural similarity search

**TABLE 1****Crystallographic data collection and refinement statistics**

Datasets were collected at the Proxima 1 beamline (Synchrotron Soleil, Saint-Aubin, France).

	Kce native orthorhombic	Kce native tetragonal <sup>a</sup>	Kce-KAH	Kce-acetoacetate
<b>Data collection</b>				
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P4 <sub>3</sub> 2 <sub>1</sub> 2	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P4 <sub>3</sub> 2 <sub>1</sub> 2
Wavelength (Å)	0.9724	0.8856	0.9801	0.8856
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	74.87, 100.82, 147.43	102.19, 102.19, 101.07	98.50, 98.48, 103.18	102.10, 102.10, 100.74
$\alpha$ , $\beta$ , $\gamma$ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	19.5–1.59 (1.68–1.59)	41.6–1.28 (1.35–1.28)	34.4–1.75 (1.84–1.75)	41.6–1.40 (1.48–1.40)
Unique reflections	147,835	133,532	101,452	104,015
<i>I</i> / $\sigma$ ( <i>I</i> ) <sup>b</sup>	16.5 (3.0)	12.4 (3.3)	16.3 (2.9)	12.5 (3.0)
<i>R</i> <sub>merge</sub>	0.066 (0.501)	0.083 (0.466)	0.061 (0.434)	0.074 (0.504)
Completeness (%) <sup>b</sup>	99.3 (95.7)	97.6 (89.6)	95.2 (92.0)	92.6 (88.5)
Redundancy <sup>b</sup>	6.7 (4.5)	7.0 (5.8)	3.6 (3.3)	7.2 (6.8)
<b>Refinement</b>				
Used reflections	140,296	126,786	96,199	98,582
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> (%)	15.7/17.4	14.3/16.9	15.0/17.4	16.7/18.6
No. of atoms				
Protein	8432	4309	8528	4238
Ligands/ions	16	9	45	24
Solvent	976	585	872	498
<i>B</i> -factors (Å <sup>2</sup> )				
Protein	21.5	17.2	19.9	20.0
Ligands/ions	30.4	16.6	33.3	22.1
Solvent	34.0	26.5	30.6	28.3
r.m.s. deviations				
Bond lengths (Å)	0.010	0.012	0.010	0.010
Bond angles (°)	1.00	1.47	0.99	1.06
PDB entry	2Y7D	2Y7E	2Y7F	2Y7G

<sup>a</sup> A brief description of the Kce native tetragonal crystalline form is provided in the [supplemental material](#).<sup>b</sup> Highest resolution shell is shown in parentheses.

**FIGURE 1. Overview of the Kce structure and the overall reaction.** *A*, reaction scheme. The different colors are designed to reflect the origin of carbon atoms in the 3-aminobutyryl-CoA and acetoacetate products, respectively, as determined by <sup>14</sup>C-tracer experiments by Barker and co-workers on homologous enzymes (8–10). *B*, ribbon diagram of the Kce homotetramer. *C*, top view of one monomer. Additional topology elements with respect to the canonical ( $\beta/\alpha$ )<sub>8</sub> TIM barrel fold are depicted in *dark red*, whereas the Zn<sup>2+</sup> ion is depicted as a *gray sphere*. The arrow points to the mobile loop connecting  $\beta$ 3 to  $\alpha$ 3. *D*, electrostatic surface representation of the active site region, with the entrances to the Zn<sup>2+</sup> crevice indicated by the *yellow arrows*. The  $\beta$ 3– $\alpha$ 3 loop, which closes over the active site upon KAH binding, is depicted in *red* and indicated by a *black arrow* as in *C*.

showed significant matches (*C* $\alpha$  r.m.s. deviation  $\sim$ 1.2 Å) to seven structures of proteins of unknown function, all composed of the Pfam database DUF849 domain (PDB entries 3NO5, 3CHV, 3FA5,

3LOT, 3E02, 3E49, and 3C6C) and showing a conserved divalent metal ion and its coordinating environment. However, no complexes with substrates or products, nor biochemical data



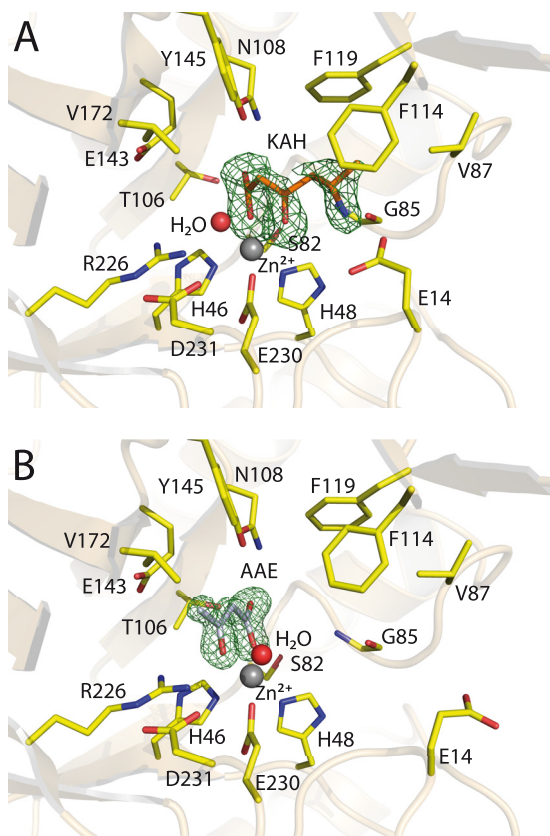


FIGURE 2. **View of the Kce active site in complex with either KAH (A) or acetoacetate (AAE) (B).** In both views, the  $\sigma_A$ -weighted difference electron density map ( $mF_o - DF_c$ ), calculated before each ligand was added to the model and contoured at the  $3\sigma$  level, is shown in the form of a green mesh. Neighboring side chains are depicted, as well as the water molecules likely to be involved in the reaction (Fig. 3).

about these enzymes, were available, and no mechanistic clues on catalysis could thus be inferred from a simple structural comparison.

**Complexes with 3-Keto-5-amino-hexanoate and Acetoacetate**—To get insights into the catalytic mechanism, we crystallized Kce in complex with either the substrate (*S*)-3-keto-5-amino-hexanoate ((*S*)-KAH) or the product acetoacetate in two different space groups (Table 1). Both molecules were found to bind at full occupancy and could be modeled unambiguously in the active site (Fig. 2). The binding of KAH did not promote significant overall structural changes ( $C\alpha$  r.m.s. deviation  $\sim 0.9$  Å) as the main chain displacements are essentially limited to the  $\beta 3$ - $\alpha 3$  loop and the  $\alpha 3$  helix (residues 82–100). In particular, the  $\beta 3$ - $\alpha 3$  loop, floppy in the native enzyme, is well structured upon the bound KAH closing one of the two entrances to the active site and leaving only one access to the  $Zn^{2+}$  crevice for the second substrate acetyl-CoA (Fig. 1D). Further, the side chains of most residues surrounding the active site pocket adopt different rotamers with respect to the unliganded enzyme, reorienting toward the bound KAH. The  $\beta$ -keto acid binds as a bidentate chelator of  $Zn^{2+}$ , with a carboxylate oxygen as well as the keto group coordinating the divalent ion at distances of 2.0 and 2.2 Å, respectively (Fig. 2A). The binding of the substrate is further stabilized by the hydrogen bonds of the carboxylate of KAH with the hydroxyl groups of Ser-82 and Thr-106 and the amide group of Asn-108, therefore preventing

the decarboxylative formation of enolate as observed in decarboxylating condensing enzymes of fatty acid and polyketide synthesis (25). The  $\delta$ -amino group of KAH interacts with the main chain carbonyl of Gly-85 and the carboxyl of Glu-14 (Fig. 2A). These interactions play a key role in the proper positioning of the *S*-ligand, suggesting a *S*-stereopreference of the enzyme. Indeed, the kinetic data show a moderate stereospecificity, the relative activity with (*R*)-KAH being 40% of the activity obtained with (*S*)-KAH (initial rates obtained with 600  $\mu M$  KAH at a saturating concentration of acetyl-CoA). In addition to Gly-85, the  $\beta 3$ - $\alpha 3$  loop also contributes to the binding of the substrate through Val-87, whose side chain makes hydrophobic interactions with the terminal methyl group of KAH. The side chains of Val-87, Phe-114, and Phe-119 define a small hydrophobic cleft in which the methyl group of KAH is inserted (Fig. 2A). A well ordered water molecule could be placed approximately halfway between the  $NH_2$  of the fully conserved Arg-226 and the carbon 2 of KAH, the predicted site of C–C cleavage in the course of the reaction, to which it is positioned at an average distance of 2.9 Å, suggesting a possible role in catalysis (see below).

The structure of Kce in complex with the reaction product acetoacetate, although obtained from crystals belonging to a different space group, shows a virtually identical enzyme fold ( $C\alpha$  r.m.s. deviation = 0.3 Å) when compared with the complex with KAH. Acetoacetate also binds to the  $Zn^{2+}$  ion as a bidentate ligand, with the carboxylate group superimposable to that of KAH, making the same interactions with Ser-82, Thr-106, and Asn-108 (Fig. 2B). However, the orientation of the molecule is approximately perpendicular to the six-carbon chain of KAH because the ketonyl oxygen of acetoacetate coordinates the metal from the Arg-226 side, with which it interacts firmly (Fig. 2B). This oxygen atom actually displaces the water molecule observed, in both the native enzyme and the KAH complex, to make a hydrogen bond with the guanidinium group of Arg-226 (see above), whereas the methyl group of acetoacetate makes van der Waals contacts with Val-172. Another water molecule now completes the  $Zn^{2+}$  coordination sphere in the position otherwise occupied by the  $\beta$ -keto oxygen of KAH.

**Reaction Mechanism and Structural Models**—From the two snapshot structures, (*S*)-KAH-Kce and acetoacetate-Kce, and considering the primary results on the origin of carbon atoms of reaction products through tracer experiments (10), we suggest a potential reaction mechanism, depicted in Fig. 3. We hypothesize that the condensation catalyzed by Kce is a  $Zn^{2+}$ -dependent, Claisen-like mechanism (25) that proceeds through the formation of four intermediates: an enolate derived from (*S*)-KAH (here named I1), a first tetrahedral high-energy intermediate (I2), a second tetrahedral high-energy intermediate (I3), and eventually an enolate form of acetoacetate (I4). Computational modeling of the predicted reaction intermediates shows compatible positioning in the active site for all of them (Fig. 4), in agreement with our hypothesis. Indeed, it has been reported that high-energy intermediates are more likely to adopt catalytically competent poses in molecular docking than ground-state intermediates (26, 27). Moreover, this kind of Claisen-like mechanism is fully compatible with the catalyzed reaction being reversible, as reported earlier from other Kce homologs (8–10).

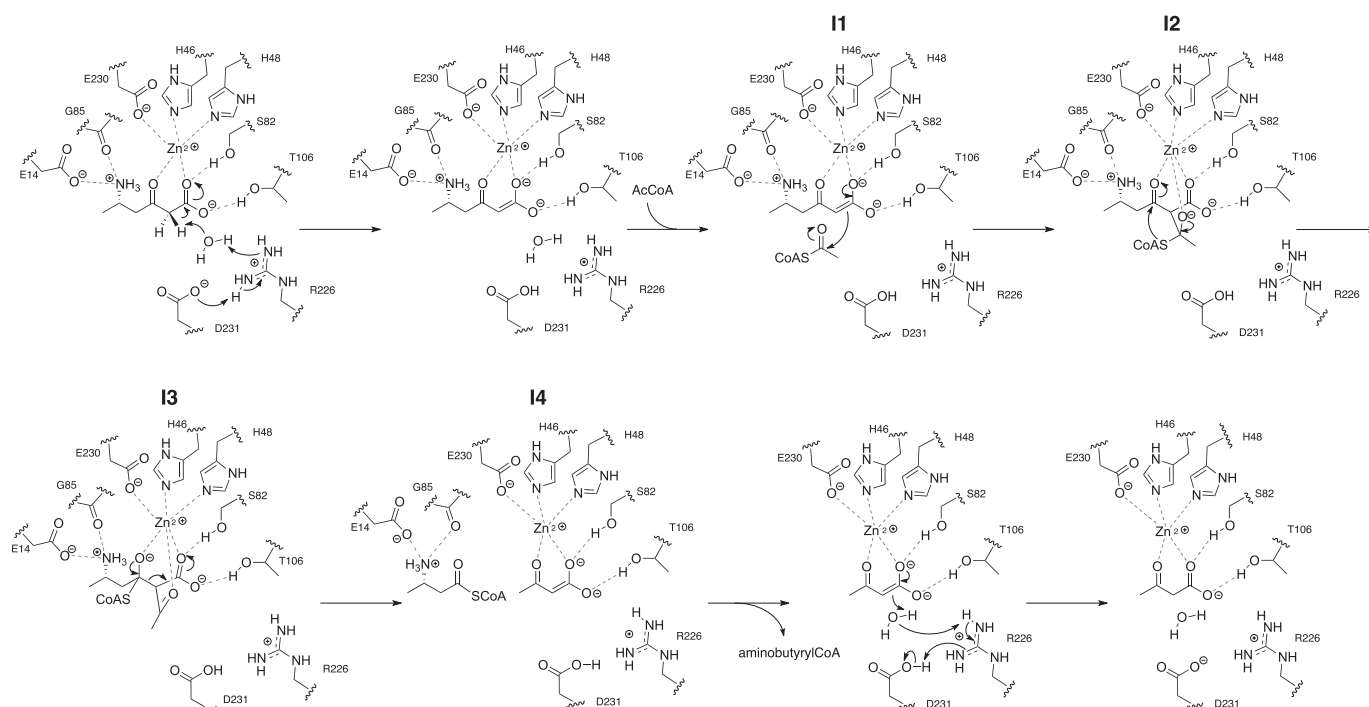


FIGURE 3. **Schematic representation of the course of the Kce reaction.** I1, I2, I3, and I4 refer to the postulated reaction intermediates (see “Results and Discussion” for details).

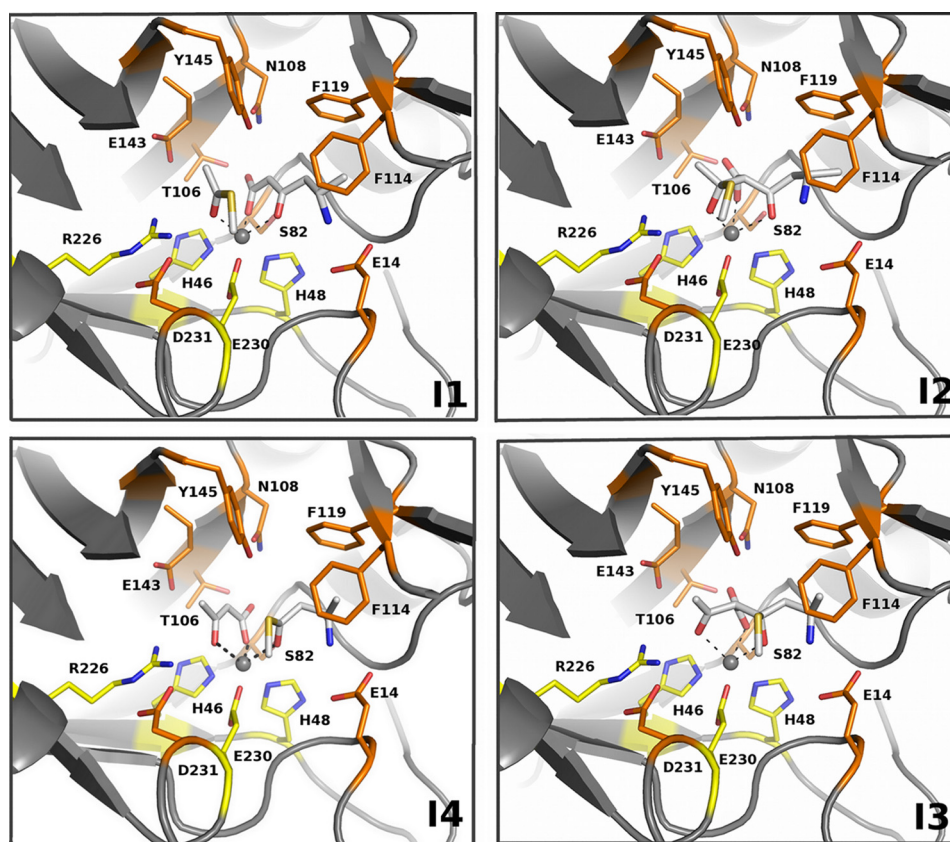


FIGURE 4. **View of the Kce active site at different states of the reaction, as predicted by molecular docking of the reaction intermediates I1, I2, I3, and I4 (illustrated in Fig. 3).** The CoA pantetheine chain, which is orientated toward the outside of the pocket, has been omitted for clarity.

The crystal structure of the KAH-Kce complex shows a strong salt bridge connecting Asp-231 to Arg-226 ( $d = 2.9 \text{ \AA}$ ), with the guanidinium group from the latter also being engaged

into a hydrogen bond to a nearby water molecule ( $d = 2.9 \text{ \AA}$ ), which, in turn, is close enough to the C2 of KAH ( $d = 2.9 \text{ \AA}$ ) to abstract the proton in a series of sequential acid/base exchanges

TABLE 2

Steady-state kinetic analysis of Kce and point mutants

ND, not determined.

Mutant	$K_m$ (S)-KAH	$K_m$ AcCoA	$k_{cat}$	$k_{cat}/K_m$ ((S)-KAH)	$k_{cat}/K_m$ (AcCoA)
	$mM$	$mM$	$s^{-1}$	$mM^{-1} \cdot s^{-1}$	$mM^{-1} \cdot s^{-1}$
S82G	$1.181 \pm 0.209$	$0.046 \pm 0.010$	$0.13 \pm 0.03$	0.1	2.8
E143G	$0.921 \pm 0.100$	$0.024 \pm 0.002$	$0.37 \pm 0.01$	0.4	15.4
E143Q	$0.442 \pm 0.061$	$0.012 \pm 0.002$	$0.32 \pm 0.05$	0.7	26.7
R226G	ND	ND	ND	—	—
D231G	ND	ND	ND	—	—
Wild type	$0.208 \pm 0.032$	$0.012 \pm 0.003$	$2.69 \pm 0.15$	12.9	224.2

initiated by Asp-231 (28). We propose that the condensation starts by the abstraction of the pro-S proton from C2 of KAH, mediated by a catalytic water molecule activated by the Asp-231-Arg-226 charge relay dyad. The key role of this Asp-Arg dyad in catalysis is strengthened by the observation that its disruption, illustrated by the Arg-226 or Asp-231 to glycine variants, leads to a catalytically inactive enzyme (Table 2). According to our hypothesis, the stability of the resulting negatively charged KAH enolate (I1) is ensured by interactions with the  $Zn^{2+}$  cation and through hydrogen bonds with Thr-106 and Ser-82, as observed in the crystal structure of the Kce-KAH complex and further supported by the docking model (Fig. 4, I1). Steady-state kinetic data are also consistent with this scheme as the replacement of Ser-82 by a glycine significantly increases the  $K_m$  for KAH and decreases the  $k_{cat}$  (Table 2). Although initial attempts of docking acetyl-CoA onto the crystallographic structure of the KAH-Kce complex failed, suggesting that either KAH adopts a different conformation in the presence of acetyl-CoA or the active site undergoes further conformational changes to accommodate both substrates, a perfect match could be obtained with the two reactive entities, *i.e.* I1 and acetyl-CoA, in the binding pocket (Fig. 4). According to this model, the orientation of the pantetheine chain of acetyl-CoA is parallel to (S)-KAH (supplemental Fig. S2A). The carbonyl carbon of acetyl-CoA is within van der Waals contact range to both C1 and C2 of KAH (respective distances are 3.4 and 3.1 Å) and is therefore well placed to react with the deprotonated carbon C2 of KAH (Figs. 3 and 4, I1). Nucleophilic addition of enolate I1 to the thioester carbon of the acetyl-CoA would lead to a tetrahedral oxyanion intermediate I2 strongly coordinated to the  $Zn^{2+}$  ion (Fig. 3). As suggested by docking calculations, the KAH-derived part of I2 is perfectly superimposed on the position of the substrate in the KAH-Kce crystallographic structure (Fig. 4, I2). The oxyanion from the acetyl moiety from acetyl-CoA takes exactly the same position as the oxygen atom of the well ordered water molecule and is stabilized by a hydrogen bond with Arg-226, whereas the three oxygen atoms (*i.e.* the oxyanion from acetyl-CoA and the carboxylic and the  $\beta$ -keto oxygens from KAH) form a tripod-like structure with the peak being the  $Zn^{2+}$  ion (Fig. 3, I2). In this conformation, the sulfur atom of acetyl-CoA makes polar interactions with Tyr-145 and van der Waals contacts with Phe-114, Val-172, and Arg-226, whereas the methyl group of the acetyl moiety is stabilized through van der Waals contacts with Glu-143 and Val-172. The short distance between the sulfur atom of the CoA moiety and the C3 keto atom of the former KAH (3.2 Å in the docking model), as well as the angle between the two reactive centers ( $\sim 100^\circ$ , close to the Bürgi-Dunitz trajectory

(29)), suggest that the intermediate may collapse through an intramolecular SCoA nucleophilic addition to the C3  $\beta$ -keto group (Fig. 3). According to molecular docking, the newly obtained tetrahedral oxyanion intermediate I3 is also expected to coordinate the  $Zn^{2+}$  ion in a similar fashion as I2 (Fig. 4). However, the sulfur atom of acetyl-CoA is here less strongly stabilized by Tyr-145 ( $d = 3.5$  Å) and is mainly held by weak van der Waals contact with Phe-114. This conformation is consistent with an intramolecular retro-Claisen reaction that would release 3-aminobutyryl-CoA as the leaving group and lead to the formation of acetoacetate in the enolate form, bound to  $Zn^{2+}$  ion by a bidentate interaction mediated by the carboxyl and  $\beta$ -keto groups, as well as by hydrogen bonds to Ser-82 and Thr-106 (Figs. 3 and 4). It is worth noting that the predicted conformation of acetoacetate at the end of the catalytic process coincides with the binding mode observed in the high-resolution crystal structure of the enzyme-product complex (Fig. 2B) and is consistent with the origin of carbon atoms in acetoacetate as determined by tracer experiments with  $^{14}C$  made by Barker and co-workers (8–10). The protonation of the enolic form of acetoacetate, possibly by the action of another water molecule taking part in a proton relay with the Asp-231-Arg-226 dyad (in a reverted process with respect to the deprotonation of KAH), would finally release the acetoacetate (Figs. 3 and 4).

In all docking models (I1, I2, I3, and I4), the adenosine moiety from the CoA ester is situated inside or near the shallow pocket flanked by the side chains of Arg-209 and Lys-237 (supplemental Fig. S2). In contrast, it should be noted that the pantetheine chain is modeled differently in the different intermediates. Indeed, CoA can adopt widely varying conformations in different enzymes (30), making its mode of binding difficult to predict (31).

Although the first attempts at modeling acetyl-CoA onto the crystallographic structure of KAH-Kce suggested a role of Glu-143 in stabilizing this substrate in the active site, activity measurements of the Glu-143 to Gly and Gln mutants invalidated this hypothesis as none of these mutations notably altered the  $K_m$  of the enzyme for acetyl-CoA. Rather, their main effect is to affect the  $k_{cat}$ , and to a lesser extent, the  $K_m$  for KAH (Table 2). Although unexpected, this effect may be explained by the interactions that, in the apo enzyme, Glu-143 makes with Arg-226 ( $d = 3.0$  Å), Asn-108 ( $d = 3.4$  Å), and Glu-141 ( $d = 2.5$  Å) (supplemental Fig. S3A). As crystallographic data revealed that Asn-108 interacts with (S)-KAH (Fig. 2A), Glu-143 variants may alter or suppress the hydrogen bond with the carbamide group of Asn-108, and in turn, slightly lower the stabilization of KAH in the active site, leading to an increased  $K_m$ . Moreover,



Glu-143 is involved in the network of interactions that hold the catalytic water molecule in position, as observed in both the apo Kce structure and the complex with KAH (supplemental Fig. S3), in which it adopts a different rotamer upon substrate binding and structuring of the  $\beta 3$ - $\alpha 3$  loop (see above). Substitutions of this residue are thus likely to alter the network that holds the key water molecule, decreasing the overall catalytic efficiency.

**Concluding Remarks**—The wide functional diversity of TIM barrel domains, which at present account for almost 9% of the whole PDB, is well known. The vast majority of these proteins are enzymes, spread ubiquitously in all kingdoms of life and most often involved in metabolic pathways (32). However, despite the already known large chemical variety of reactions catalyzed by this fold, new types of reactions or novel catalytic mechanisms performed by TIM barrel enzymes could still have escaped identification. Indeed, we provide here strong evidence pointing to Kce as the first representative of a new subclass of TIM barrel  $\text{Zn}^{2+}$ -dependent condensing enzymes. Similarly to canonical class II aldolases, which share with Kce both the presence of the zinc cation and the same  $(\beta/\alpha)_8$  barrel fold,  $\text{Zn}^{2+}$  is crucial to chelate both the substrate and the nascent acetoacetate product, as well as to stabilize their enolic forms, but the overall reaction course is mechanistically closer to a carbon-carbon Claisen condensation. Our model is fully consistent with earlier literature reports about the origin of carbon atoms in the products and the fact that no covalent acyl-enzyme intermediate could ever be observed (8–10). Further research would reveal whether or not Kce represents a member of a whole new class of enzymes catalyzing carbon-carbon condensations involving CoA esters and small organic  $\beta$ -keto acids.

**Acknowledgments**—We are grateful to C. Pellé and C. Lechaplais (CEA Genoscope) for excellent technical assistance and to A. Haouz and P. Weber (PF6, Institut Pasteur) for performing robot-driven crystallization trials. We also thank ESRF and Synchrotron SOLEIL for provision of synchrotron radiation facilities, and their respective staffs for assistance in using the beamlines.

## REFERENCES

- Bateman, A., Coghill, P., and Finn, R. D. (2010) *Acta Crystallogr. F Struct. Biol. Cryst. Commun.* **66**, 1148–1152
- Dayhoff, M. O., Schwartz, R., and Orcutt, B. C. (1978) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., ed) pp. 345–352, National Biomedical Research Foundation, Washington, D.C.
- Eddy, S. R. (1996) *Curr. Opin. Struct. Biol.* **6**, 361–365
- Osterman, A., and Overbeek, R. (2003) *Curr. Opin. Chem. Biol.* **7**, 238–251
- Zhang, C., and Kim, S. H. (2003) *Curr. Opin. Chem. Biol.* **7**, 28–32
- Jaroszewski, L., Li, Z., Krishna, S. S., Bakolitsa, C., Wooley, J., Deacon, A. M., Wilson, I. A., and Godzik, A. (2009) *PLoS Biol.* **7**, e1000205
- Kreimeyer, A., Perret, A., Lechaplais, C., Vallenet, D., Médigue, C., Salanoubat, M., and Weissenbach, J. (2007) *J. Biol. Chem.* **282**, 7191–7197
- Barker, H. A., Kahn, J. M., and Chew, S. (1980) *J. Bacteriol.* **143**, 1165–1170
- Barker, H. A., Kahn, J. M., and Hedrick, L. (1982) *J. Bacteriol.* **152**, 201–207
- Yorifuji, T., Jeng, I. M., and Barker, H. A. (1977) *J. Biol. Chem.* **252**, 20–31
- Kabsch, W. (2010) *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132
- Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763
- Sheldrick, G. M. (2008) *Acta Crystallogr. A* **64**, 112–122
- Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. D Biol. Crystallogr.* **55**, 849–861
- Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501
- Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255
- Vagin, A., and Teplyakov, A. (1997) *J. Appl. Crystallogr.* **30**, 1022–1025
- Blanc, F. A., Roversi, P., Vonnrhein, C., Flensburg, C., Lea, S. M., and Bricogne, G. (2004) *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2210–2221
- Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21
- Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004) *Nucleic Acids Res.* **32**, W665–667
- Davis, F. A., Zhang, Y., Andemichael, Y., Fang, T., Fanelli, D. L., and Zhang, H. (1999) *J. Org. Chem.* **64**, 1403–1406
- Trott, O., and Olson, A. J. (2010) *J. Comput. Chem.* **31**, 455–461
- Hinsen, K. (2000) *J. Comput. Chem.* **21**, 79–85
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) *J. Comput. Chem.* **25**, 1605–1612
- Heath, R. J., and Rock, C. O. (2002) *Nat. Prod. Rep.* **19**, 581–596
- Hermann, J. C., Ghanem, E., Li, Y., Raushel, F. M., Irwin, J. J., and Shoichet, B. K. (2006) *J. Am. Chem. Soc.* **128**, 15882–15891
- Hermann, J. C., Marti-Arbona, R., Fedorov, A. A., Fedorov, E., Almo, S. C., Shoichet, B. K., and Raushel, F. M. (2007) *Nature* **448**, 775–779
- Guillén Schlippe, Y. V., and Hedstrom, L. (2005) *Arch. Biochem. Biophys.* **433**, 266–278
- Burgi, H. B., Dunitz, J. D., Lehn, J. M., and Wipff, G. (1974) *Tetrahedron* **30**, 1563–1572
- Anstrom, D. M., Kallio, K., and Remington, S. J. (2003) *Protein Sci.* **12**, 1822–1832
- Koon, N., Squire, C. J., and Baker, E. N. (2004) *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8295–8300
- Nagano, N., Orengo, C. A., and Thornton, J. M. (2002) *J. Mol. Biol.* **321**, 741–765